REVIEW ARTICLE

The molecular population genetics of regulatory genes

MICHAEL D. PURUGGANAN

Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC 27695 USA

Abstract

Regulatory loci, which may encode both *trans* acting proteins as well as *cis* acting promoter regions, are crucial components of an organism's genetic architecture. Although evolution of these regulatory loci is believed to underlie the evolution of numerous adaptive traits, there is little information on natural variation of these genes. Recent molecular population genetic studies, however, have provided insights into the extent of natural variation at regulatory genes, the evolutionary forces that shape them and the phenotypic effects of molecular regulatory variants. These recent analyses suggest that it may be possible to study the molecular evolutionary ecology of regulatory diversification by examining both the extent and patterning of regulatory gene diversity, the phenotypic effects of molecular variation at these loci and their ecological consequences.

Keywords: evolution of development, phenotypic variation, polymorphism, promoters, QTL, transcriptional activators

Received 23 February 2000; revision received 11 April 2000; accepted 11 April 2000

Introduction

Regulatory genes represent a class of loci that control the expression of other genes (Patel 1994; McSteen & Hake 1998). These genes play central roles in the genetic architecture of eukaryotic development because they guide developmental trajectories and specify organismal morphologies (McSteen & Hake 1998). Developmental genetic studies indicate that various classes of regulatory loci control the differentiation of diverse structures, and it has been suggested that interspecies morphological diversity arises from evolution of these developmental genes (Carroll 1995; Palopoli & Patel 1996; Doebley & Lukens 1998; Purugganan 1998). These loci may also play important roles in physiological adaptation by controlling the temporal and spatial expression of structural enzymatic loci. There has been increasing interest in understanding the evolution of genes that regulate developmental processes, many of which have been shown to encode sequence-specific, DNA-binding transcriptional activators or cell–cell signalling molecules (Meyerowitz 1999). Efforts have also been underway to study the contribution of *cis*-acting regulatory sequences such as promoter regions in shaping regulatory

evolution (Dickinson 1988; Doebley & Lukens 1998). Our knowledge of the evolutionary forces that shape the diversification of regulatory systems, however, remains extremely limited, and this has hampered our ability to systematically investigate the role of regulatory gene evolution on organismal divergence and adaptation.

Studies of regulatory gene evolution have focused primarily on macroevolutionary patterns of gene diversification. Recent phylogenetic analyses of several regulatory gene families reveal that most developmental systems evolve by duplication and divergence of paralogous loci (Purugganan *et al*. 1995; Zhang & Nei 1996; Purugganan 1998). Moreover, some regulatory genes appear to evolve faster than typical structural genes (Purugganan & Wessler 1994; Purugganan 1998; Ting *et al*. 1998) and these elevated molecular evolutionary rates may be associated with developmental alterations, morphological diversification, physiological adaptations and even speciation.

It is clear, however, that sequence evolution of loci at the macroevolutionary level arises from molecular variation that occurs within distinct species. All molecular variation must have its origins at the level of populations, where evolutionary forces play themselves out to determine the fates of alternate alleles (Palopoli & Patel 1996; Purugganan 1998; Arthur 1999). Investigation of the evolution of these loci both within populations and between

Correspondence: Michael D. Purugganan. Fax: (919) 515 3355; E-mail: michaelp@unity.ncsu.edu

closely related species is required for a comprehensive understanding of the evolutionary dynamics of these control genes. A molecular population genetic approach to the study of regulatory gene evolution provides a framework for assessing how population history, breeding system and selection affect variation at these genetic loci, and for delineating mechanisms that lead to evolutionary diversification. This approach should prove particularly effective when comparing the partitioning of variation and gene genealogies across species boundaries, allowing us to connect microevolutionary forces within species to macroevolutionary change between species.

The study of the molecular population genetics of regulatory genes revolves around three questions: Is there within-species variation in regulatory genes? What are the evolutionary forces that shape genetic variation at these control loci? And finally, does molecular variation at these regulatory genes result in phenotypic variation within populations and between species? By addressing these questions, we can establish the role of regulatory genes in evolutionary diversification as well as providing links between phenotypic variation in organismal development and physiology on the one hand, and molecular evolution on the other.

Recent studies on the molecular population genetics of regulatory loci have begun to shed light on some of the evolutionary forces that shape developmental gene structure, and their possible links to morphological evolution and, possibly, adaptive diversification. Work in *Arabidopsis thaliana*, *Zea mays*, *Brassica oleracea* and *Drosophila melanogaster* suggests that regulatory genes can harbour significant levels of molecular variation, and that both selective (Walthour & Schaeffer 1994; Wang *et al*. 1999; Purugganan *et al*. 2000), recombinational (Wang *et al*. 1999) and demographic (Purugganan & Suddith 1999) forces interact to pattern variation at developmental loci. These studies are beginning to provide the empirical scaffolding for understanding the mechanisms that underlie the evolution of these pivotal genes.

Molecular variation and evolutionary forces at regulatory loci

Although there has been great interest in the role of regulatory genes in evolution, it was unclear, until fairly recently, whether regulatory loci harboured sufficient genetic variation within species to serve as the raw material for evolutionary diversification. The dramatic, often pleiotropic phenotypes of mutant alleles of many regulatory loci suggested that these genes would be under strong stabilizing selection and thus possess little variation, particularly at the protein level. Moreover, the strong conservation of regulatory proteins across highly divergent taxa were noted in comparative molecular studies

(McGinnis *et al*. 1984), which reinforced the suggestion that stabilizing selection was the dominant force that constrained the evolution of these regulatory genes.

In the last decade, the levels and patterning of molecular diversity of various developmental regulatory genes have been examined and, in general, the levels of diversity at these regulatory genes do not appear to differ from other types of loci in the genomes of these species. The presence of polymorphisms at the population level indicates that these genes possess molecular variation that, in principle, could be the targets of selective forces. However, there are indications that many regulatory loci may harbour reduced levels of diversity compared to structural genes. In a compilation of data for sequence variation at 22 nuclear genes in *Drosophila melanogaster*, for example, regulatory genes appear to posses less than half of the diversity of structural genes (Moriyama & Powell 1996). For X-linked loci, the estimate of the mean proportion of nucleotide differences (π) is 0.0015 for regulatory genes vs. 0.0036 for structural genes (Purugganan 1998). Additionally, the mean estimate of nucleotide diversity for autosomal regulatory genes (0.019) is less than that observed for structural loci (0.0049) (Purugganan 1998).

Several studies also suggest reduced variation in levels of protein polymorphism at some of these developmental genes. In *D. melanogaster*, the *transformer* gene is an RNA binding protein necessary for correct sexual differentiation of somatic cells in females. In a study of variation among 10 alleles of this locus, no amino acid polymorphisms were observed (Walthour & Schaeffer 1994). A similar pattern is seen for the three *D. melanogaster Ras*-related genes (*Dras1–3*) (Gasperini & Gibson 1999). *Dras1* is essential for axial pattern formation, segmentation and organogenesis in developing embryos, while *Dras3* (also known as *Roughened*) is involved in photoreceptor determination of eye imaginal discs. The function of *Dras2* is as yet unknown. Like other *D. melanogaster* regulatory genes, these three developmental loci display reduced levels of nucleotide variation and show no amino acid polymorphism in a sample of 27 alleles (Gasperini & Gibson 1999).

The forces that act to reduce variation at regulatory genes can differ, depending upon the gene in question. Reduced molecular diversity at specific loci may arise from the action of adaptive or positive selection on a gene. The low levels of nucleotide variation observed for the *terminal ear1*, a putative RNA binding protein (White & Doebley 1999) and the *C1* basic helix-loop-helix transcription factor locus (Hanson *et al* 1996) in *Zea mays* ssp. *mays*, has been attributed to selection. Selection has also been implicated in low variation observed at the floral regulatory gene *BoCAULIFLOWER* (*BoCAL*) in *B. oleracea* ssp. *botrytis* and ssp. *italica* (Purugganan *et al*. 2000). In these loci, the reduced diversity may be associated with positive selection during the domestication of crop species.

In other instances, the lowered molecular diversity of different genes may arise as a consequence of their genetic context within the genome. For example, the segment polarity gene *cubitus interruptus* of *D. melanogaster*, which encodes a zinc-finger DNA-binding protein, and the *prune* GTPase activating protein locus are among the least diverse loci in the *Drosophila* genome, with estimates of π at silent sites being 0.00 (Berry *et al.* 1991; Simmons *et al.* 1994). The low levels of diversity for these regulatory loci, however, are associated with their location in chromosomal regions of low recombination (Begun & Aquadro 1992), and it is likely that this reduced variation may arise in part from the presence of positively selected sites at close linkage to these regulatory genes. Both *cubitus interruptus* and *prune* indicate how the position of developmental loci along the chromosome does impact rates of molecular evolution, and illustrates the importance of genomic context in shaping the levels of variation at regulatory genes.

In contrast, increased protein polymorphism is observed at three floral developmental genes in *Arabidopsis thaliana*, a wild weed in the mustard oil (Brassicaceae) family. The *CAULIFLOWER*, *APETALA3* and *PISTILLATA* genes encode MADS-box DNA-binding transcriptional activators which control different aspects of flower development (Riechmann & Meyerowitz 1997). The *AP3* and *PI* genes are both necessary for petal and stamen differentiation in the developing *Arabidopsis* flower. *CAULIFLOWER* is a recent (~30 Ma) duplicate of *APETALA1* and both share partially redundant functions in specifying the developmental identity of flower primordia in inflorescence stems. In all three of these genes there is an excess of within-species replacement nucleotide polymorphisms compared to synonymous coding region variation. In a sample of 17 alleles of the *CAL* locus, for example, 16 of the 21 coding region polymorphisms are nonsynonymous and change the amino acid sequence of the encoded transcriptional activators (Purugganan & Suddith 1998). A similar pattern of elevated within-species replacement polymorphisms are observed in the *AP3* and *PI* genes (Purugganan & Suddith 1999). The McDonald-Kreitman test of protein evolution indicates that these increases in replacement polymorphisms across all three floral developmental genes are significant when compared to the relative levels of replacement and synonymous differences between these loci in *A. thaliana* and its close relative *A. lyrata*.

The pattern of elevated within-species protein polymorphism in these three plant regulatory genes is shared with several other loci in the *Arabidopsis* genome, and may reflect the impact of demographic forces in shaping variation at these developmental genes (Purugganan & Suddith 1999). Although selective arguments (including local adaptation) can be invoked to explain this increase in within-species regulatory protein variation, it is more likely that this diversity reflects slightly deleterious polymorphisms that persist due to reduced effective population sizes in this inbreeding plant or recent population expansion in this species (Purugganan & Suddith 1999). Both the reduction in diversity in some *Drosophila* regulatory loci and the increase in protein polymorphism in *Arabidopsis* floral developmental genes illustrate how factors acting either outside these loci (such as genetic hitchhiking in chromosomal regions of reduced recombination) or demographic forces (population substructuring, expansion) affect the levels and patterning of diversity at developmental loci — a situation that may have implications on how we examine the evolutionary genetics of developmental pattern diversification within and between species.

Selective forces at regulatory genes

The extent and patterning of diversity of genes are shaped by a myriad of evolutionary forces acting on organisms and their genomes. Specifically, it is of interest to examine the extent to which diversification of regulatory genes both within and between species is governed by either neutral drift or selection, the two main forces that determine evolution at the molecular level. Understanding the nature of the selective forces that impact these loci should provide further insights into mechanisms by which molecular evolution at these control genes play a role in organismal diversification.

Variation at many of the regulatory genes studied, either within populations or between closely related species, appears to be largely neutral. Analyses of the *Drosophila melanogaster runt* locus, a primary pair rule gene encoding a transcription factor (Labate *et al.* 1999), and *bride-of-sevenless* (*boss*), which produces a EGF-like ligand protein involved in photoreceptor determination (Ayala & Hartl 1993), are both examples of regulatory genes whose patterns of diversity are consistent with neutral evolution. Most other *Drosophila* loci studied, including *decapentaplegic* (*dpp*) (Richter *et al.* 1997) and *Dras1–3* (Gasperini & Gibson 1999), also appear to be evolving neutrally, although the latter genes also appear to exhibit strong constraints limiting protein sequence change. *Drosophila* regulatory genes are similar in this respect to many of their structural gene counterparts, in which neutral evolution remains the dominant force that determines the levels and patterns of molecular diversity.

Variation at the *Zea mays* ssp. *mays terminal ear1* (*te1*) (White & Doebley 1999) and *C1* genes (Hanson *et al.* 1996) have been implicated as targets of selection by early New World farmers during the domestication of maize. The *te1* gene encodes a protein that partially controls male inflorescence development, while the C1 *myb*-like protein partially regulates the pigmentation of maize kernels.

Tajima D   = -2.5088 (P < 0.05)*
Fu & Li D* = -3.6731 (P < 0.05)*

Tajima D   = -0.7070 (n.s.)
Fu & Li D* = 1.1645 (n.s.)
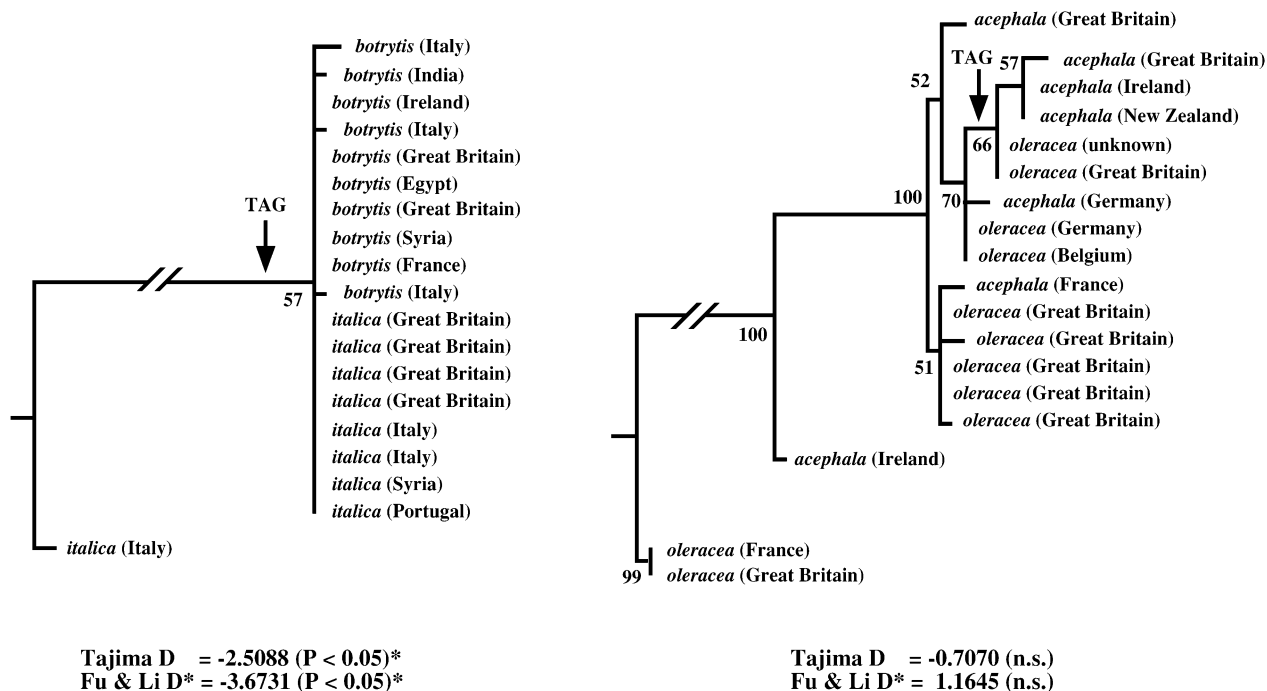
Brassica oleracea ssp. *botrytis* and ssp. *italica*   *Brassica oleracea* ssp. *oleracea* and ssp. *acephala*

Fig. 1 Intraspecific gene genealogies of the *BoCAL* gene in different *Brassica oleracea* subspecies. The gene genealogy of alleles found in domesticated subspecies showing variation in inflorescence morphologies (left) is depicted separately from those subspecies that do not show evolutionary change in reproductive structures (right). The low variation in the *BoCAL* gene for *B. oleracea* ssp. *botrytis* (cauliflower) and ssp. *italica* (broccoli), as well as results of the Tajima, and Fu & Li tests for selection, indicate that the evolution of these two domesticated subspecies was accompanied by positive selection at this regulatory locus. (Purugganan *et al.* 2000).

Although statistical tests of neutrality fail to reject the hypothesis of neutral evolution, the reduced intraspecific diversity at these two genes compared to other maize loci may have arisen from positive selection during domestication (Hanson *et al.* 1996; White & Doebley 1999). Interestingly, the haplotypes of the *C1* gene in *Zea mays* ssp. *mays* and its wild teosinte relatives form two classes, and the domestication of maize appears to be associated with selection against one of these haplotype classes (Hanson *et al.* 1996).

The work in maize highlights the utility of using domesticated crop species to study the role of regulatory gene evolution in organismal diversification. Indeed, unequivocal examples of selection on regulatory genes have been observed at another maize developmental gene (the *teosinte branched1* locus discussed in the next section) as well as the *Brassica oleracea BoCAL* gene. Mutations at *BoCAL*, which encodes a MADS-box transcriptional activator involved in floral meristem development, is associated with the cauliflower head phenotypes in Brassicaceae species (Kempin *et al.* 1995). Two subspecies of the domesticated vegetable crop *Brassica oleracea* [ssp. *botrytis* (cauliflower) and ssp. *italica* (broccoli)] are characterized by the evolutionary modification of the inflore-

scence into large dense structures. The evolution of these altered morphologies is associated with the near-fixation of G → T transversion in these domesticated subspecies (Purugganan *et al.* 2000) (see Fig. 1). This polymorphism results in the replacement of a glutamic acid with a nonsense (stop) codon, and haplotypes bearing this nonsense mutation produce a truncated MADS-box protein. Variation at the *BoCAL* locus is reduced in *B. oleracea* ssp. *botrytis* and ssp. *italica* compared to other domesticated and wild *B. oleracea* subspecies, and tests of selection indicate that the pattern of variation at this regulatory gene is consistent with a recent selective sweep in domesticated cauliflower and broccoli (Purugganan *et al.* 2000) (see Fig. 1).

There is evidence that several *D. melanogaster* regulatory genes have also been subject to adaptive selection during their recent evolution. Two examples illustrate the role of adaptive selection in the divergence of *Drosophila* genes between closely related species. The *transformer* gene, involved in fly sex determination, shows little within-species protein variation while exhibiting large between-species divergence (Walthour & Schaeffer 1994). Within *D. melanogaster*, there are only two exon polymorphisms at the *tra* locus, none of which is a replacement polymorphism. In contrast, there are 16 replacement and 22

synonymous fixed differences at the *tra* gene between *D. melanogaster* and *D. simulans*. HKA tests of selection indicate that this reduction in coding sequence variation within *D. melanogaster* is significant when compared to levels of between-species divergence (Walthour & Schaeffer 1994).

Evidence for positive selection is also seen in the *OdysseusH* (*OdsH*) homeodomain gene between *D. mauritiana* and *D. simulans* (Ting *et al.* 1998). This gene, which is found in a chromosomal region associated with reproductive isolation between several *Drosophila* species, shows very rapid evolution in protein sequence at the generally conserved homeodomain DNA-binding region of the encoded transcriptional activator. The pattern of evolution at *OdsH* between these two sibling species is interpreted as evidence for positive selection at this regulatory locus within the last 500 000 years (Ting *et al.* 1998). There is no data for intraspecific variation at this locus, and it would be interesting to see whether the patterns of diversity within *OdsH* in different *Drosophila* taxa also displays the footprint of natural selection associated with reproductive isolation between species.

Variation in promoter regions

The molecular population genetics of *cis*-acting regulatory sequences, such as those found in promoter regions and occasionally 3′ gene regions and introns, are less well-studied than the *trans*-acting regulatory genes that encode DNA binding transcription factors or signalling proteins. There has always been intense interest in examining the evolution of promoter-encoded *cis* regulatory sequences, as it is widely believed that it is molecular diversification at these control sequences that are of pivotal importance in organismal evolution (Dickinson 1988). Indeed, there have been suggestions that even for *trans*-acting regulatory genes, it is the promoters, rather than protein coding regions, that are the primary targets of adaptive evolution (Doebley & Lukens 1998).

A clear example of adaptive evolution in promoter sequences is demonstrated by the maize *tb1* gene (Doebley *et al.* 1997). The *tb1* gene is involved in the evolution of shoot apical dominance in maize, and is related to the *cycloidea* gene of *Antirrhinum majus* (snapdragons). Both these genes possess a putative nuclear localization signal and share a novel protein domain (called the TCP domain) with other proteins that are believed to function as transcriptional activators (Cubas *et al.* 1999). Molecular population genetic analysis of the transcription unit of the *tb1* locus in both domesticated maize and related wild teosinte subspecies show no evidence of selection in the coding region of this regulatory locus (Wang *et al.* 1999). Intraspecific diversity in the promoter sequence of this gene, however, is reduced in the domesticated *Zea mays*
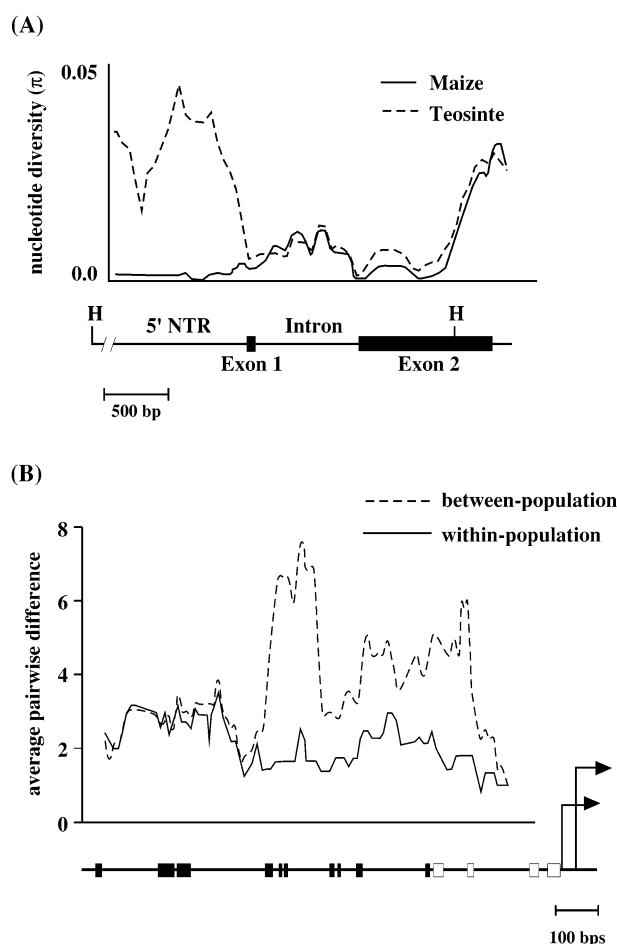

Fig. 2 Variation in promoter sequences. (A) Intraspecific variation in the *tb1* promoter sequence in domesticated maize and its wild teosinte relative. The variation is plotted with respect to gene, indicated below the plot. The maize *tb1* allele shows reduced levels of variation in the promoter but not coding region. Adapted from Wang *et al.* 1999. (B) Variation in *Ldh-B* promoter alleles within- and between northern and southern populations of *Fundulus heteroclitus*. Variation across the promoter is low between populations, but a central region of elevated between-population divergence is evident. The position of the major transcription start sites are indicated by the bent arrows. Solid and open boxes show the position of indel and microsatellite variants between alleles. Adapted from Schulte *et al.* 1997.

ssp. *mays* compared to variation within wild teosinte subspecies (see Fig. 2A). Tests of selection indicate that this reduction in polymorphism in the domesticated group is significant and consistent with a recent adaptive sweep at or near the promoter of *tb1* (Wang *et al.* 1999). Interestingly, the effects of this selective sweep are confined to the 5′ region of the gene and does not extend into the coding region of *tb1*; presumably, recombination in maize is sufficiently strong to break the linkage between the promoter of this regulatory gene and its coding region. Given typical recombination rates for maize, the selection

coefficient at the *tb1* promoter during this selective sweep was estimated at between 0.04 and 0.08 — fairly strong selection but possibly typical of crop domestication events (Wang *et al.* 1999).

The action of selection at promoter sequences (albeit in an enzymatic, not regulatory gene) has also been documented for the *Ldh-B* promoter of the teleost fish *Fundulus heteroclitus* (Schulte *et al.* 1997). Analysis of a 1-kb region immediately 5′ of the *Ldh-B* housekeeping enzyme gene reveals the presence of two major allelic classes which appear to correspond to differentiation between northern and southern populations of this fish along the North American Atlantic seaboard. Although the diversity of the *Ldh-B* promoter sequence is low, there is a central region that shows strong divergence between Maine and Florida *F. heteroclitus* populations (see Fig. 2B). The divergence in sequence between northern and southern *Ldh-B* promoters is significant using the HKA test, suggesting that adaptive divergence is responsible for the differentiation of *cis*-acting regulatory elements at this housekeeping gene between these geographical populations. Significantly, this promoter divergence is correlated with differences in *Ldh-B* transcription levels, and functional deletion analyses indicates the presence of a transcriptional repressor element in southern but not northern *Ldh-B* promoter alleles. The differences in *Ldh-B* transcription levels attributed to various naturally occurring promoters has in turn been implicated in between-population variations in hatching times, developmental rates, swimming performance and differential mortality at elevated temperatures (Crawford & Powers 1992).

Although both the *tb1* and *Ldh-B* promoters exhibit evidence of positive selection, other promoter sequences appear to evolve neutrally. One example is the promoter of the *Drosophila even-skipped* (*eve*) gene, which plays a key role in embryonic segmentation. The *eve* promoter contains a 671-bp enhancer (the MSE) that partially controls expression of the *eve* gene in a spatial manner during embryonic development. Molecular population genetic analysis of this enhancer region within *D. melanogaster* and *D. simulans* reveals nucleotide and insertion/deletion variation throughout the promoter, and the pattern of diversity suggests that this region is evolving neutrally (Ludwig & Kreitman 1995). Interestingly, portions of the *eve* promoter, including the MSE, appears to be functionally constrained; the rates of evolution for these promoter regions are approximately one-third that observed for the *eve* intron. Moreover, the presence of mutations that eliminate transcription factor binding sites in this promoter, including *Kruppel* regulatory protein binding sites, suggests that binding site redundancy in these promoters buffer this *cis*-acting regulatory sequence against the phenotypic consequences of mutational change (Ludwig & Kreitman 1995).

Indeed, recent analyses indicate there is high turnover in binding sites within the *eve* promoter between different *Drosophila* species, although functional analyses reveal that promoters from different species still result in conserved patterns of *eve* expression (Ludwig *et al.* 2000). These evolutionary functional assays indicate that stabilizing selection has been important in maintaining patterns of *eve* expression despite evolutionary changes in promoter structure. A model of enhancer evolution has been proposed which suggests that weakly selected mutations in regulatory elements will be present in natural populations and available for positive selection, possibly in the context of multiple compensatory mutaions across the promoter sequence (Ludwig *et al.* 2000).

Although these three examples highlight the diverse ways that promoter sequences can evolve within and between species, we still know relatively little about how these *cis*-acting regulatory sequences evolve. More information will be necessary if we are to ascertain the relative contributions of variation in promoter vs. regulatory protein coding regions in adaptive diversification. There is also a need to correlate the evolution of *trans*-acting regulatory proteins to their *cis*-acting promoter sequence targets; together, this will provide a fuller picture of the molecular evolution of gene regulatory systems.

Genotype to phenotype — does molecular variation matter?

Although the data are still meager, we now know that regulatory genes can harbour substantial levels of diversity at the molecular level, and that various evolutionary forces, including positive selection, determine both the levels and patterning of the observed variation present at these control loci. A major challenge in molecular evolution and ecology in the coming years is to forge the final causal links between variation at the molecular level and the evolution of phenotypes encoded by these genes. We must figure out whether the variation we observe at the molecular level actually matters to the organism, and whether the changes we observe within populations among genes correlates with phenotypic evolution. If we can make the connection between molecules and phenotypes, then we open whole new fields of enquiry that ultimately will lead to a comprehensive understanding of the historical, genetic and ecological components of organismal adaptation. There are now several studies that attempt to bridge the chasm between molecular and phenotypic diversity, in work ranging from investigations of wild populations to domesticated plant species.

Domestic plant species provide excellent models to study and test hypotheses on relationships between molecular variation at regulatory loci and phenotypic variation within and between species (Doebley 1993). The
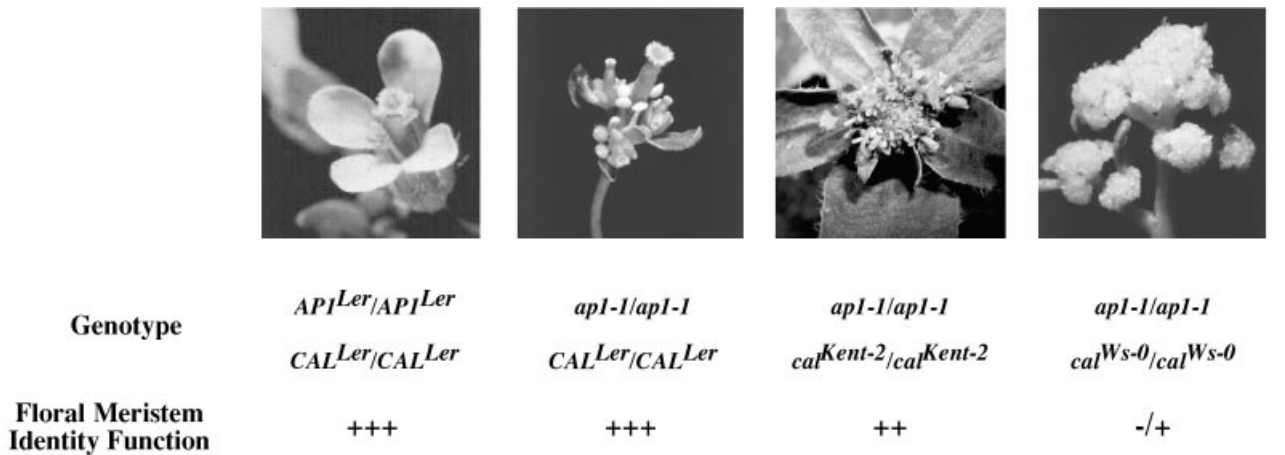
Genotype	API^{Ler}/API^{Ler} CAL^{Ler}/CAL^{Ler}	$ap1-1/ap1-1$ CAL^{Ler}/CAL^{Ler}	$ap1-1/ap1-1$ $cal^{Kent-2}/cal^{Kent-2}$	$ap1-1/ap1-1$ cal^{Ws-0}/cal^{Ws-0}
Floral Meristem Identity Function	+++	+++	++	-/+

Fig. 3 Functional differences between naturally occurring *CAULIFLOWER* alleles in *Arabidopsis thaliana*. Differences between *CAL* alleles in their ability to form floral meristems (as indicated by the plus signs) is evident in an *apetala1–1* null mutant background. Much of the protein variation that distinguishes the three *CAL* alleles map to the C-terminal domain of the encoded protein. The phenotype of the *ap1–1 cal^{WS–0}* plants can be rescued by complementation in transgenic plants (Kempin *et al*. 1995).

domestication of crop species is invariably accompanied by evolutionary changes in suites of structural traits that differentiate cultivated species from their wild relatives, or even between various crop subspecies. Thus, crop species have been widely regarded as providing some of the best and most dramatic examples of the degree to which plant morphologies evolve under selection pressures (Gottlieb 1984; Doebley 1993).

The utility of looking at domestic crop species to inform our understanding of molecular regulatory basis of evolutionary change is highlighted in two examples discussed previously — the *teosinte branched1* gene of *Zea mays* and *BoCAL* of *Brassica oleracea*. In the former, the evolution of domesticated *Zea mays* ssp. *mays* from the wild *Zea mays* ssp. *parviglumis* and ssp. *mexicana* is accompanied by an increase in apical dominance. Quantitative trait locus (QTL) analysis indicates that differences in basal shoot (tiller) formation between these subspecies are controlled in part by a gene (or genes) on the long arm of chromosome 1 (Doebley & Stec 1991). This chromosomal region encompasses the *tb1* locus, whose mutant phenotype in maize is reminiscent of the morphology of the wild teosinte ancestor. Transposon tagging in maize led to the cloning of the *tb1* locus, and expression studies indicated that the maize *tb1* allele expressed the gene at higher levels than the wild teosinte allele in a maize inbred background (Doebley *et al*. 1997). Consistent with the difference in transcript levels is the finding, as discussed previously, that the *tb1* promoter has undergone a recent selective sweep in *Zea mays* ssp. *mays* but not in wild *Zea mays* subspecies (Doebley *et al*. 1997).

It is not clear from the *tb1* analysis precisely what changes in the promoter region are responsible for the evolutionary variation in *tb1* activity between maize and

wild teosinte (Wang *et al*. 1999). In contrast, studies of the *BoCAL* MADS-box transcription factor gene in domesticated *Brassica oleracea* pinpoints a clear candidate for a polymorphism associated with morphological variation within *B. oleracea*. In this instance, the evolution of the cauliflower (and possibly the broccoli) phenotype is associated in part with the presence of a stop codon in exon 5 that results in the production of a truncated protein (Kempin *et al*. 1995), consistent with genetic studies in the *CAULIFLOWER* orthologue in the related crucifer *Arabidopsis thaliana*. Evidence in *B. oleracea*, both from molecular population and developmental genetic analysis, indicates that the evolution of the unique reproductive morphologies in *B. oleracea* ssp. *botrytis* can be traced, in part, to at least one single nucleotide change in a floral regulatory locus.

Although studies in domesticated species are revealing, the strong selection pressures that accompany domestication events raise questions as to how far we can draw parallels between regulatory gene evolution in crop species and evolution under natural selection in wild populations (Coyne & Lande 1985). In wild species, however, evidence suggests that molecular polymorphisms are clearly associated with phenotypic variation, and may have the potential to play roles in organismal evolution. Genetic studies of the *A. thaliana CAULIFLOWER* gene, for example, have shown that naturally occurring alleles in this floral homeotic locus are functionally distinguishable (Purugganan & Suddith 1998). In this gene, replacement polymorphisms in exon 7 are associated with a differential ability of alleles to specify floral meristems in different *Arabidopsis* ecotypes (Bowman *et al*. 1993; Kempin *et al*. 1995; Purugganan & Suddith 1998) (see Fig. 3), and this difference may be associated with natural variation in inflorescence branch number (Bowman *et al*. 1993). In another study, variation
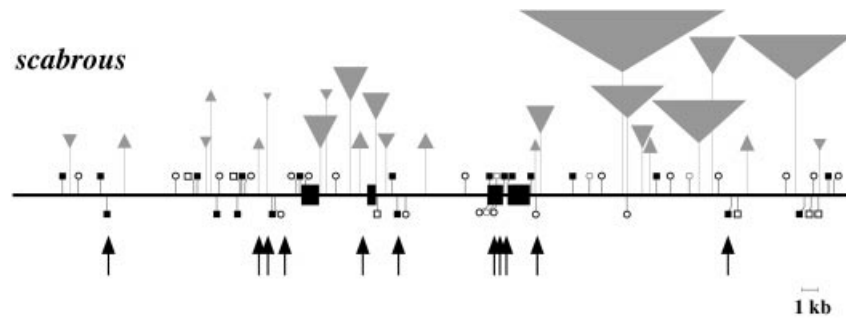
Fig. 4 Molecular diversity at the *scabrous* locus associated with variation in *Drosophila* bristle number. The *scabrous* exons large solid boxes. Variation was assayed by restriction site mapping: solid squares, open squares and open circles denote *Pst*I, *Bam*HI and *Eco*RI sites, respectively. Monomorphic and polymorphic restriction sites in a sample of 47 haplotypes are shown above and below the gene, respectively. Insertion/deletions are also indicated. Eleven polymorphisms significantly associated with bristle number variation are identified by the arrows. Adapted from Lai *et al.* 1994.

in the *Drosophila melanogaster Ultrabithorax* locus, a homeotic gene involved in segmental identity, is associated with naturally occurring variation in the ether-induced bithorax phenocopy implicated in the phenomenon of genetic assimilation (Gibson & Hogness 1996). Variation in *Ubx* may ultimately lead to differences in other aspects of morphology between species; indeed, changes at the *Ubx* locus between *D. melanogaster*, *D. simulans* and *D. virilis*, particularly at the *cis*-acting promoter regions, are associated with differences in trichome patterning in the fly leg (Stern 1998).

Analyses of natural populations of *D. melanogaster* have also uncovered associations between quantitative variation in sensory bristle number and various genes that regulate the development of peripheral nervous system (Mackay 1996). The *achaete scute* complex (*ASC*), for example, defines proneural regions in the fly and mutations within this complex reduce the size of these regions and the number of sensory bristles in *D. melanogaster*. Naturally occurring variation in bristle number has been correlated with polymorphisms near the *scute α, β* and *γ* basic helix-loop-helix (bHLH) genes of the *achaete scute* complex (Mackay & Langley 1990; Long *et al.* 2000). A small deletion polymorphism near *sc α* accounts for 25% of the genetic variation in sternopleural bristle number contributed by the X chromosome, while a 3.4-kb insertion between *sc β* and *γ* accounts for 22% of the X chromosome-associated variation in female abdominal bristle number (Long *et al.* 2000). The presence of these polymorphisms at intermediate frequency in natural fly populations suggest that they are maintained neutrally or by balancing selection. Moreover, large insertions in the *ASC* are associated with natural reductions in bristle numbers, and their low frequencies in the population indicate that the evolutionary dynamics of these insertion polymorphisms are largely governed by deleterious mutation-selection balance.

Quantitative variation in bristle number has also been associated with natural variation in the neurogenic locus *scabrous*, which encodes a secreted glycoprotein with a structure similar to growth factors and is important in lateral inhibition of the developing nervous system. Eleven polymorphic sites were detected that account for 32% of abdominal and 21% of sternopleural bristle number in *D. melanogaster* populations (Lai *et al.* 1994) (see Fig. 4). These polymorphisms are found throughout the gene, including 5′ and 3′ regions as well as introns, and exist in intermediate frequencies in a wild population. Polymorphisms in two introns of *Delta*, a neurogenic locus that encodes a ligand in the *Notch* signalling pathway, also account for a significant fraction of natural genetic variation in fly bristle numbers (Long *et al.* 1998).

Molecular polymorphisms in these three regulatory genes are clearly associated with natural phenotypic variation of a quantitative trait; what remains to be seen is whether these polymorphisms are causal (or merely linked to the causal) mutations that lead to variation in bristle number. It is clear, however, that the field of view in the hunt for the polymorphisms that define naturally occurring phenotypic variation in bristle number has been considerably narrowed, and it is only a matter of time until geneticists pinpoint the precise polymorphisms in these and other candidate regulatory genes that are the direct causes of diversity in this quantitative trait.

Finally, one of the more intriguing links between molecular and phenotypic diversity at regulatory loci is seen in the toadflax *Linaria vulgaris*. In this plant species, some individuals in wild British populations display peloric variants, possessing radially symmetric as opposed to the normal bilaterally symmetric zygomorphic flowers (Cubas *et al.* 1999). This natural variation in floral symmetry was first described by Linnaeus, and recent work indicates it is correlated with changes in the *cycloidea* (*Lcyc*) floral regulatory gene of *Linaria*. Interestingly enough, however,

the phenotypic variation observed does not appear to arise from mutational polymorphisms in *Lcyc*, but in methylation variants of the gene sequence in peloric plants (Cubas *et al.* 1999). We do not know the degree to which epigenetic changes such as methylation may influence stable, heritable phenotypic variation in natural populations. The results from the *Lcyc* locus, as well as other recent reports on the importance of methylation changes in gene activity, suggests that these epigenetic mechanisms may play some role in generating phenotypic diversity in populations. It will be intriguing to examine the frequency of heritable epigenetic changes in natural populations and the impact they may have on the action of selection as well as long-term ecological adaptation and evolutionary change.

A molecular evolutionary ecology of adaptation

Despite the relatively small number of studies on the population diversity of regulatory loci, we already have preliminary answers to the three questions we posed earlier. We now know that despite the specter of stabilizing selection, regulatory genes continue to harbour significant levels of molecular diversity. We also know that various evolutionary forces — from neutral-drift to adaptive selection — have worked to pattern the variation we observe at these control loci And finally, we can correlate, in some instances, variation of the molecules with phenotypic evolution at the organismal level.

We still need to know more. In particular, we have to expand the number of population-level studies of regulatory genes in order to obtain a more comprehensive picture of the levels and patterning of diversity at these loci. Moreover, we need to systematically investigate the molecular population genetics not only of individual loci but whole developmental pathways — whether those that regulate embryonic development in *Drosophila melanogaster* or flower development in *Arabidopsis thaliana* or other regulatory networks. Indeed, molecular population geneticists need to better exploit the wealth of knowledge on the genes and gene interactions that make up organismal regulatory pathways. Several questions remain to be addressed: (i) are there differences in the levels of variation (and constraint) between genes that act earlier or later in regulatory pathways? (ii) Can we discern patterns of molecular coevolution between genes that directly interact with one another, or between transcriptional activators and their target promoter elements? And (iii) are there trends in the modes and strengths of evolutionary forces that operate on different molecular components of a regulatory hierarchy?

An important aspect of future research will be to assay, whenever possible, the functional significance of the polymorphisms we observe at the molecular level. To relate evolutionary change with functional differentiation, we need to investigate how variation in regulatory proteins within species translates into transcriptional activator efficiency, or how promoter sequence diversity results in changes in target gene transcriptional levels. Several studies have already begun to explore these issues (Schulte *et al.* 1997), but continued effort is required if we are to better understand the mechanistic consequences, at the molecular and cellular level, of evolution in *trans*- and *cis*-acting regulatory loci.

These questions are related to perhaps the most important avenue of exploration in this area, which is to examine the direct links between variation at these regulatory genes and evolution at the phenotypic level. It is already clear from genetic analysis that mutations at several regulatory genes are accompanied by both striking qualitative as well as quantitative variation in numerous developmental and physiological traits. It must be established in most instances whether the results of genetic mutant screens reflect wild reality, and whether the corresponding natural variation in phenotypic traits we observe in wild populations can be traced to specific polymorphisms at the molecular level. The search for the genetic links to naturally occurring phenotypic variation will be aided significantly by the availability of complete genome sequences in several evolutionarily and ecologically relevant model species (such as *D. melanogaster* and *A. thaliana*); whole genome sequences provide opportunities to quickly establish genetic linkages between molecular polymorphisms and phenotypic diversity.

If we succeed, we can then begin to investigate the evolution of adaptations by examining, in one overarching research programme, the ecological forces that act on specific traits, the variation in observed developmental or physiological regulation associated with phenotypes, the molecular genetic mechanisms, and the evolutionary forces that lie behind a specific change at the molecular level. Only then can we say we possess a more comprehensive understanding of the possible role of regulatory genes in organismal diversification. This is the challenge of an emerging molecular evolutionary ecology of adaptation.

Acknowledgements

References

Arthur W (1999) Variable segment number in centipedes: Population genetics meets evolutionary developmental biology. *Evolution of Development*, **1**, 62–69.

Ayala FJ, Hartl DL (1993) Molecular drift of the *bride-of-sevenless (boss)* gene in *Drosophila*. *Molecular Biology and Evolution*, **10**, 1030–1040.

Begun DJ, Aquadro DF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. Melanogaster*. *Nature*, **356**, 519–520.

Berry AJ, Ajioka JW, Kreitman M (1991) Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics*, **129**, 1111–1117.

Bowman JL, Alvarez J, Weigel D, Meyerowitz EM, Smyth DR (1993) Control of flower development in *Arabidopsis thaliana* by *AP1* and interacting genes. *Development*, **119**, 721–743.

Carroll SB (1995) Homeotic genes and the evolution of arthropods and chordates. *Naturalist*, **376**, 479–485.

Coyne JA, Lande R (1985) The genetic basis of species differences in plants. *American Naturalist*, **126**, 141–145.

Crawford DL, Powers DA (1992) Evolutionary adaptation to different thermal environments via transcriptional regulation. *Molecular Biology and Evolution*, **9**, 806–813.

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, **401**, 157–161.

Dickinson WJ (1988) On the architecture of regulatory systems: Evolutionary insights and implications. *Bioessays*, **8**, 204–208.

Doebley J (1993) Genetics, development and plant evolution. *Current Opinion in Genetics and Development*, **3**, 865–872.

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell*, **10**, 1075–1082.

Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. *Genetics*, **129**, 285–295.

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485–488.

Gasperini R, Gibson G (1999) Absence of protein polymorphism in the *Ras* genes of *Drosophila Melanogaster*. *Journal of Molecular Evolution*, **49**, 583–590.

Gibson G, Hogness DS (1996) Effect of polymorphisms in the *Drosophila* regulatory gene *Ubx* on homeotic stability. *Science*, **271**, 200–203.

Gottlieb LD (1984) Genetics and morphological evolution in plants. *American Naturalist*, **123**, 681–709.

Hanson MA, Gaut BS, Stec AO *et al.* (1996) Evolution of anthocyanin biosynthesis in maize kernels: The role of regulatory and enzymatic loci. *Genetics*, **143**, 1395–1407.

Kempin S, Savidge B, Yanofsky MF (1995) Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science*, **267**, 522–525.

Labate JA, Biermann CH, Eanes WF (1999) Nucleotide variation at the *runt* locus in *Drosophila melanogaster* and *D. simulans*. *Molecular Biology and Evolution*, **16**, 724–731.

Lai C, Lyman RF, Long AD, Langley CH, Mackay TFC (1994) Naturally occurring variation in bristle number and DNA polymorphisms at the *scabrous* locus of *Drosophila Melanogaster*. *Science*, **266**, 1697–1702.

Long AD, Lyman RF, Langley CH, Mackay TFC (1998) Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila Melanogaster*. *Genetics*, **149**, 999–1017.

Long AD, Lyman RF, Morgan AH, Langley CH, Mackay TFC (2000) Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila Melanogaster*. *Genetics*, **154**, 1255–1269.

Ludwig MZ, Kreitman M (1995) Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila*. *Molecular Biology and Evolution*, **12**, 1002–1011.

Ludwig MZ, Bergmann C, Patel NH, Kreitman M (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*, **403**, 564–567.

Mackay TFC (1996) The nature of quantitative genetic variation revisited: Lessons from *Drosophila* bristles. *Bioessays*, **18**, 113–121.

Mackay TFC, Langley CH (1990) Molecular and phenotypic variation in the *achaete-scute* region of *Drosophila melanogaster*. *Nature*, **348**, 64–66.

McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell*, **37**, 403–408.

Meyerowitz EM (1999) Plants, animals and the logic of development. *Trends in Cell Biology*, **9**, M65–M68.

McSteen P, Hake S (1998) Genetic control of plant development. *Current Opinion in Biotechnology*, **9**, 189–195.

Moriyama EN, Powell JR (1996) Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution*, **13**, 261–277.

Palopoli MF, Patel N (1996) Neo-Darwinian developmental evolution — can we bridge the gap between pattern and process? *Current Opinion in Genetics and Development*, **6**, 502–508.

Patel NH (1994) Developmental evolution — insights from studies of insect segmentation. *Science*, **266**, 581–590.

Purugganan MD (1998) The molecular evolution of development. *Bioessays*, **20**, 700–711.

Purugganan MD, Boyles AL, Suddith JI (2000) Variation and selection at the *CAULIFLOWER* floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics*, **155**, 855–862.

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF (1995) Molecular evolution of flower development: Diversification of the plant MADS-box regulatory gene family. *Genetics*, **140**, 354–356.

Purugganan MD, Suddith JI (1998) Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: Non-neutral evolution and wild variation in floral homeotic function. *Proceedings of the National Academy of Sciences of the USA*, **95**, 8130–8134.

Purugganan MD, Suddith JI (1999) Molecular population genetics of floral homeotic loci: Departures from the equilibrium-neutral model at the *AP3* and *PI* genes of *Arabidopsis thaliana*. *Genetics*, **151**, 839–848.

Purugganan MD, Wessler SR (1994) Molecular evolution of the plant *R* regulatory gene family. *Genetics*, **138**, 849–854.

Richter B, Long M, Lewontin RC, Nitasaka E (1997) Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in *Drosophila*. *Genetics*, **145**, 311–323.

Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. *Biological Chemistry*, **378**, 1079–1101.

Schulte PM, Gomez-Chiarri M, Powers DA (1997) Structural and functional differences in the promoter and 5′ flanking region of *Ldh-B* within and between populations of the teleost *Fundulus Heteroclitus*. *Genetics*, **145**, 759–769.

Simmons GM, Kwok W, Matulonis P, Venkatesh T (1994) Polymorphism and divergence at the *prune* locus in *Drosophila melanogaster* and *D Simulans*. *Molecular Biology and Evolution*, **11**, 666–671.

Stern DL (1998) A role of *Ultrabithorax* in morphological differences between *Drosophila* species. *Nature*, **396**, 463–466.

Ting CT, Tsaur SC, Wu ML, Wu CI (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science*, **282**, 1501–1504.

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature*, **398**, 236–239.

Walthour CS, Schaeffer SW (1994) Molecular population genetics of sex determination genes: The *transformer* gene of *Drosophila melanogaster*. *Genetics*, **136**, 1367–1372.

White SE, Doebley JF (1999) The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics*, **153**, 1455–1462.

Zhang J, Nei M (1996) Evolution of *Antennapedia*-class homeobox genes. *Genetics*, **142**, 295–303.